The Mann-Whitney test can be inverted to form confidence intervals. Let us consider a "shift" model: $G(x) = F(x - \Delta)$. This model says that the effect of the treatment (the Y's) is to add a constant Δ to what the response would have been with no treatment (the X's). (This is a very simple model, and we have already seen cases for which it is not appropriate.) We now derive a confidence interval for Δ. To test H_0: $F = G$, we used the statistic U_Y equal to the number of the $X_i - Y_j$ that are less than zero. To test the hypothesis that the shift parameter is Δ, we can similarly use

$$U_Y(\Delta) = \#[X_i - (Y_j - \Delta) < 0] = \#(Y_j - X_i > \Delta)$$

It can be shown that the null distribution of $U_Y(\Delta)$ is symmetric about $mn/2$:

$$P\left(U_Y(\Delta) = \frac{mn}{2} + k\right) = P\left(U_Y(\Delta) = \frac{mn}{2} - k\right)$$

for all integers k. Suppose that $k = k(\alpha)$ is such that $P(k \leq U_Y(\Delta) \leq mn - k) = 1 - \alpha$; the level α test then accepts for such $U_Y(\Delta)$. By the duality of confidence intervals and hypothesis tests, a $100(1 - \alpha)\%$ confidence interval for Δ is thus

$$C = \{\Delta \mid k \leq U_Y(\Delta) \leq mn - k\}$$

C consists of the set of values Δ for which the null hypothesis would not be rejected.

We can find an explicit form for this confidence interval. Let $D_{(1)}, D_{(2)}, \ldots, D_{(mn)}$ denote the ordered mn differences $Y_j - X_i$. We will show that

$$C = [D_{(k)}, D_{(mn-k+1)})$$

To see this, first suppose that $\Delta = D_{(k)}$. Then

$$\begin{aligned} U_Y(\Delta) &= \#(X_i - Y_j + \Delta < 0) \\ &= \#(Y_j - X_i > \Delta) \\ &= mn - k \end{aligned}$$

Similarly, if $\Delta = D_{(mn-k+1)}$,

$$\begin{aligned} U_Y(\Delta) &= \#(Y_j - X_i > \Delta) \\ &= k \end{aligned}$$

(You might find it helpful to consider the case $m = 3$, $n = 2$, $k = 2$.)

EXAMPLE **C** We return to the data on iron retention (Section 11.2.1.1). The earlier analysis using the t test rested on the assumption that the populations were normally distributed, which, in fact, seemed rather dubious. The Mann-Whitney test does not make this assumption. The sum of the ranks of the Fe^{2+} group is used as a test statistic (we could have as easily used the U statistic). The rank sum is 362. Using the normal approximation to the null distribution of the rank sum, we get a p-value of .36. Again, there is insufficient evidence to reject the null hypothesis that there is no differential retention. The 95% confidence interval for the shift between the two distributions is $(-1.6, 3.7)$, which overlaps zero substantially. Note that this interval is shorter than